

(18)

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-254494

(43)Date of publication of application : 25.09.1998

(51)Int.Cl.

G10L 3/02

G10L 3/00

G10L 9/16

(21)Application number : 09-056018

(71)Applicant : ASAHI CHEM IND CO LTD

(22)Date of filing : 11.03.1997

(72)Inventor : SHOSAKAI MAKOTO

(54) SPEECH RECOGNITION DEVICE AND METHOD THEREFOR

(57)Abstract:

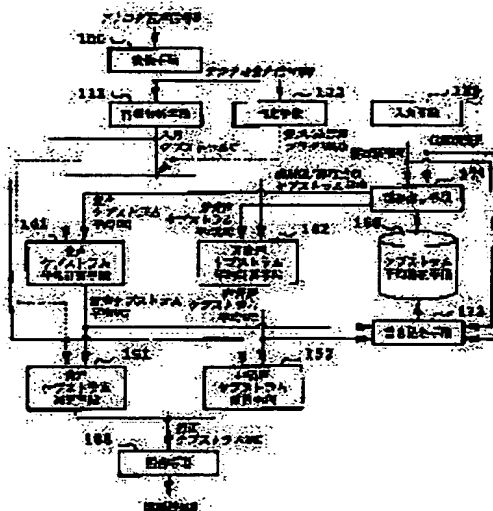
PROBLEM TO BE SOLVED: To improve accuracy of speech recognition by providing a signal processing means to correct multiplication distortion of courtroom area related to sound in an input signal to be recognized, based on a feature of the sound in a storage means designated by a designation means.

SOLUTION: A feature collection mode and an instruction of a speech recognition mode, discrimination information added to the collected features and stored, for example, discrimination numbers, descriptions, etc., and discrimination information given when taking out features used for speech recognition from a spectrum mean storage means 166 are inputted from an input means.

The speech cepstrum subtraction means 151 subtracts a speech cepstrum mean VC from an input cepstrum C.

Thus, multiplication distortions caused by difference in frequency characteristics of microphones, difference in speech characteristics due to vocal tract of speakers and vocal chords, and difference in positional

relationship between speaker and microphone can be corrected by batch processing.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's]

Best Available Copy

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平10-254494

(43) 公開日 平成10年(1998) 9月25日

(51) Int.Cl.⁶
G 1 0 L 3/02 3 0 1
3/00 5 3 1
9/16 3 0 1

F I
G 1 0 L 3/02 3 0 1 C
3/00 5 3 1 K
9/16 3 0 1 A

審査請求 未請求 請求項の数18 O L (全 11 頁)

(21) 出願番号 特願平9-56018

(22) 出願日 平成9年(1997) 3月11日

(71) 出願人 000000033

旭化成工業株式会社

大阪府大阪市北区堂島浜1丁目2番6号

(72) 発明者 庄境 誠

神奈川県厚木市棚沢221番地 旭化成工業株式会社内

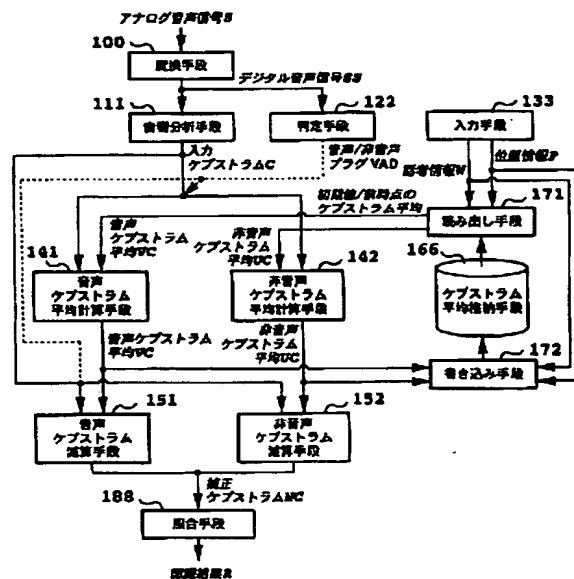
(74) 代理人 弁理士 谷 義一

(54) 【発明の名称】 音声認識装置および方法

(57) 【要約】

【課題】 音声認識精度を向上させる。

【解決手段】 話者の音声の歪みの特徴(ケブストラム平均)や特定環境での音響の歪みの特徴(ケブストラム平均)をケブストラム平均格納手段166に格納しておき、音声認識時には格納してある特徴の中の該当の歪みの特徴を使用して音声認識時に入力された音声信号の歪みを補正して音声認識する。



【特許請求の範囲】

【請求項1】 特定の音のみの入力信号から前記音の特徴を抽出する抽出手段と、

前記抽出手段から抽出された音の特徴を記憶する記憶手段と、

前記記憶手段に記憶された音の特徴の1つを指定する指定手段と、

前記指定手段により指定された前記記憶手段の音の特徴に基づき認識対象の入力信号の中の前記音についてスペクトラム領域の乗法歪みを補正する信号処理手段とを具えたことを特徴とする音声認識装置。

【請求項2】 請求項1に記載の音声認識装置において、前記抽出手段は、前記特定の音のみの入力信号に対してケプストラム分析を行い、分析の結果得られるケプストラムを平均し、当該平均化したケプストラム平均を前記音の特徴とすることを特徴とする音声認識装置。

【請求項3】 請求項1に記載の音声認識装置において、情報を入力する情報入力手段をさらに有し、当該入力された情報を前記抽出された音の特徴に付加して前記記憶手段に記憶することを特徴とする音声認識装置。

【請求項4】 請求項3に記載の音声認識装置において、前記記憶手段に記憶された情報を読み出す読み出し手段と、当該読み出した情報を表示する表示手段をさらに具えたことを特徴とする音声認識装置。

【請求項5】 請求項1に記載の音声認識装置において、前記音は非音声であることを特徴とする音声認識装置。

【請求項6】 請求項1に記載の音声認識装置において、前記音は特定話者の音声であることを特徴とする音声認識装置。

【請求項7】 請求項1に記載の音声認識装置において、前記記憶手段は、着脱自在の記録媒体であることを特徴とする音声認識装置。

【請求項8】 請求項1に記載の音声認識装置において、前記音のみの入力信号の波形が音声波形か非音声波形かを判定する判定手段と、前記記憶手段に記憶すべき音の特徴について音声か非音声かを種類指定する種類指定手段とをさらに有し、前記判定手段の判定結果に基づき、前記抽出手段は前記種類指定手段により指定された音声または非音声の音の特徴を前記特定の音のみの入力信号から抽出することを特徴とする音声認識装置。

【請求項9】 スペクトラム領域の乗法歪みの補正のための複数種類の音の特徴を予め記憶した記憶手段と、前記記憶手段に記憶された音の特徴の1つを指定する指定手段と、

前記指定手段により指定された前記記憶手段の音の特徴に基づき認識対象の入力信号の中の前記音についてスペクトラム領域の乗法歪みを補正する信号処理手段とを具えたことを特徴とする音声認識装置。

【請求項10】 音響分析回路により特定の音のみの入

力信号から前記音の特徴を抽出し、

当該抽出された音の特徴を記憶回路に記憶しておき、

当該記憶された音の特徴の1つを指定し、

当該指定された音の特徴に基づき認識対象の入力信号の中の前記音についてスペクトラム領域の乗法歪みを信号処理回路により補正することを特徴とする音声認識方法。

【請求項11】 請求項10に記載の音声認識方法において、前記音響分析回路は、前記特定の音のみの入力信号に対してケプストラム分析を行い、分析の結果得られるケプストラムを平均し、当該平均化したケプストラム平均を前記音の特徴とすることを特徴とする音声認識方法。

【請求項12】 請求項10に記載の音声認識方法において、情報を入力する情報入力回路をさらに有し、当該入力された情報を前記抽出された音の特徴に付加して前記記憶回路に記憶することを特徴とする音声認識方法。

【請求項13】 請求項12に記載の音声認識方法において、前記記憶回路に記憶された情報を読み出す読み出し、表示することを特徴とする音声認識方法。

【請求項14】 請求項10に記載の音声認識方法において、前記音は非音声であることを特徴とする音声認識方法。

【請求項15】 請求項10に記載の音声認識方法において、前記音は特定話者の音声であることを特徴とする音声認識方法。

【請求項16】 請求項10に記載の音声認識方法において、前記記憶回路は、着脱自在の記録媒体であることを特徴とする音声認識方法。

【請求項17】 請求項10に記載の音声認識方法において、前記音のみの入力信号の波形が音声波形か非音声波形かを判定回路により判定し、前記記憶回路に記憶すべき音の特徴について音声か非音声かを種類指定し、前記判定回路の判定結果に基づき、種類指定された音声または非音声の音の特徴を前記特定の音のみの入力信号から抽出することを特徴とする音声認識方法。

【請求項18】 スペクトラム領域の乗法歪みの補正のための複数種類の前記音の特徴を予め記憶回路に記憶しておき、

該記憶回路に記憶された音の特徴の1つを指定し、

当該指定された音の特徴に基づき認識対象の入力信号の中の前記音についてスペクトラム領域の乗法歪みを信号処理回路により補正することを特徴とする音声認識方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、異なる環境で入力された音声を認識する音声認識装置および方法に関する。

【0002】

【従来の技術】音声認識技術として、音響的特徴を記述するパラメータとして、ケプストラムを含むパラメータを使用し、標準パターンの照合の方法として、動的計画法(DP:Dynamic Programming)や隠れマルコフモデル法(HMM:Hidden Markov Model)を使用することが広く知られている。ここで、ケプストラムは、短時間スペクトル包絡の対数の逆フーリエ変換として定義される。

【0003】動的計画法や隠れマルコフモデル法の技術については、例えば、中川聖一著「確率モデルによる音声認識」((社) 電子情報通信学会) に詳述されている。いずれの方法においても、予め認識対象の語彙の特徴(ケプストラムなどの音響特徴を表現するパラメータ)とその識別情報、たとえば、語彙を表す文字列を標準パターンとして用意する。音声認識時にはマイクロホンから入力された音声の特徴と上記標準パターンの特徴と照合し、最も確からしい特徴の識別情報を選択して音声認識結果として出力する方法が一般的である。音声認識装置の実際の使用環境においては、音声が入口の口から発声されて、マイクロホンで収録され、音声認識処理のためのデジタルの電気信号に変換される信号処理系の経路に特有の周波数特性が存在する。

【0004】特有の周波数特性を生成する要因としてはマイクロホンの周波数特性、口からマイクロホンまでの伝達特性などがあげられる。上述の標準パターンを作成する際に使用したときの上記特有の周波数特性が異なっている場合、その差は音声信号のスペクトル領域の乗法性歪みとなって現れ、その歪みが顕著であると、著しく音声の認識性能が低下することが知られている。

【0005】このような不具合を解消するために従来ではケプストラム平均正規化法(CMN:Cepstrum Mean Normalization)が効果的であることが知られている。ケプストラム平均正規化方法では、2つの音声収録環境の間の乗法性歪みが十分に長い長さを持つ音声信号のケプストラム平均の差として得られることを利用し、音声信号の各フレーム毎に計算されるケプストラムからケプストラム平均を減算することにより、スペクトラム領域の乗法性歪みを補正する。この技術についてはJunqua and Haton著「Robustness Automatic Speech Recognition」(Kluwer Academic Publishers)のpp250-254に記述されている。

【0006】また、入力信号が音声フレームか非音声フレームかの判定を行って、音声フレームおよび非音声フレームそれぞれのケプストラム平均を求め、入力信号についてのケプストラム平均から上記音声フレームおよび非音声フレームのいずれかのケプストラム平均を減算することにより誤認識が低下することが、Huang, X at al. "Microsoft Windows

Highly Intelligent Speech Recognizer:Whisper." Proc. ICASSP, Detroit, 1995などの文献で知られている。

【0007】

【発明が解決しようとする課題】従来のケプストラム平均正規化法ではマイクロホンと話者との間の距離の変化や話者毎の声道、声帯の違いによるスペクトルの変化を考慮せず、一律にケプストラム平均を求めていたので、上記マイクロホンと話者の口との間の距離が異なったり、話者が異なると音声認識性能が低下するという解決すべき課題があった。

【0008】そこで、本発明の目的は、上述の点に鑑みて、話者とマイクロホンの位置関係や話者の相違に起因する歪みの影響を緩和して、音声認識性能を向上させることができる音声認識装置および方法を提供することにある。

【0009】

【課題を解決するための手段】このような目的を達成するために、請求項1の発明は、特定の音のみの入力信号から前記音の特徴を抽出する抽出手段と、前記抽出手段から抽出された音の特徴を記憶する記憶手段と、前記記憶手段に記憶された音の特徴の1つを指定する指定手段と、前記指定手段により指定された前記記憶手段の音の特徴に基づき認識対象の入力信号の中の前記音についてスペクトラム領域の乗法歪みを補正する信号処理手段とを具備したことを特徴とする。

【0010】請求項2の発明は、請求項1に記載の音声認識装置において、前記抽出手段は、前記特定の音のみの入力信号に対してケプストラム分析を行い、分析の結果得られるケプストラムを平均し、当該平均化したケプストラム平均を前記音の特徴とすることを特徴とする。

【0011】請求項3の発明は、請求項1に記載の音声認識装置において、情報を入力する情報入力手段をさらに有し、当該入力された情報を前記抽出された音の特徴に付加して前記記憶手段に記憶することを特徴とする。

【0012】請求項4の発明は、請求項3に記載の音声認識装置において、前記記憶手段に記憶された情報を読み出す読み出し手段と、当該読み出した情報を表示する表示手段をさらに具備したことを特徴とする。

【0013】請求項5の発明は、請求項1に記載の音声認識装置において、前記音は非音声であることを特徴とする。

【0014】請求項6の発明は、請求項1に記載の音声認識装置において、前記音は特定話者の音声であることを特徴とする。

【0015】請求項7の発明は、請求項1に記載の音声認識装置において、前記記憶手段は、着脱自在の記録媒体であることを特徴とする。

【0016】請求項8の発明は、請求項1に記載の音声

認識装置において、前記音のみの入力信号の波形が音声波形か非音声波形かを判定する判定手段と、前記記憶手段に記憶すべき音の特徴について音声か非音声かを種類指定する種類指定手段とをさらに有し、前記判定手段の判定結果に基づき、前記抽出手段は前記種類指定手段により指定された音声または非音声の音の特徴を前記特定の音のみの入力信号から抽出することを特徴とする。

【0017】請求項9の発明は、スペクトラム領域の乗法歪みの補正のための複数種類の音の特徴を予め記憶した記憶手段と、前記記憶手段に記憶された音の特徴の1つを指定する指定手段と、前記指定手段により指定された前記記憶手段の音の特徴に基づき認識対象の入力信号の中の前記音についてスペクトラム領域の乗法歪みを補正する信号処理手段とを具えたことを特徴とする。

【0018】請求項10の発明は、音響分析回路により特定の音のみの入力信号から前記音の特徴を抽出し、当該抽出された音の特徴を記憶回路に記憶しておき、当該記憶された音の特徴の1つを指定し、当該指定された音の特徴に基づき認識対象の入力信号の中の前記音についてスペクトラム領域の乗法歪みを信号処理回路により補正することを特徴とする。

【0019】請求項11の発明は、請求項10に記載の音声認識方法において、前記音響分析回路は、前記特定の音のみの入力信号に対してケプストラム分析を行い、分析の結果得られるケプストラムを平均し、当該平均化したケプストラム平均を前記音の特徴とすることを特徴とする。

【0020】請求項12の発明は、請求項10に記載の音声認識方法において、情報を入力する情報入力回路をさらに有し、当該入力された情報を前記抽出された音の特徴に付加して前記記憶回路に記憶することを特徴とする。

【0021】請求項13の発明は、請求項12に記載の音声認識方法において、前記記憶回路に記憶された情報を読み出す読み出し、表示することを特徴とする。

【0022】請求項14の発明は、請求項10に記載の音声認識方法において、前記音は非音声であることを特徴とする。

【0023】請求項15の発明は、請求項10に記載の音声認識方法において、前記音は特定話者の音声であることを特徴とする。

【0024】請求項16の発明は、請求項10に記載の音声認識方法において、前記記憶回路は、着脱自在の記録媒体であることを特徴とする。

【0025】請求項17の発明は、請求項10に記載の音声認識方法において、前記音のみの入力信号の波形が音声波形か非音声波形かを判定回路により判定し、前記記憶回路に記憶すべき音の特徴について音声か非音声かを種類指定し、前記判定回路の判定結果に基づき、種類指定された音声または非音声の音の特徴を前記特定の音

のみの入力信号から抽出することを特徴とする。

【0026】請求項18の発明は、音響分析回路により特定の音のみの入力信号から前記音の特徴を抽出し、スペクトラム領域の乗法歪みの補正のための複数種類の前記音の特徴を予め記憶回路に記憶しておき、当該記憶回路に記憶された音の特徴の1つを指定し、当該指定された音の特徴に基づき認識対象の入力信号の中の前記音についてスペクトラム領域の乗法歪みを信号処理回路により補正することを特徴とする。

【0027】

【発明の実施の形態】以下、図面を参照して本発明の実施の形態を詳細に説明する。

【0028】(第1の実施形態)図1は第1の実施の形態のシステム構成を示す。図1において、変換手段100は不図示のマイクロホンから入力されるアナログ信号(音声が入力されていなくてもアナログ音声信号Sと称する)をデジタルプロセッサやCPU等の情報処理プロセッサが実行可能なデジタル信号(以下、デジタル音声信号SSと称す)に変換する。

【0029】以下の手段は上記情報処理プロセッサがソフトウェアプログラムを実行することにより実現される。音響分析手段111は、デジタル音声信号SSに対して一定周期(音声フレーム単位)でケプストラム分析を行い、その分析結果として得られるケプストラム(入力ケプストラムCと称する)を出力する。判定手段122はデジタル音声信号SSの波形が図2の(a)に示すような人間の音声の波形か、非音声かを判定する。その判定結果は音声の場合、オン、非音声の場合オフのフラグVADで表される。この判定方法については古井貞熙著「デジタル音声処理」(東海大学出版会)に記載されている方法などを使用するとよい。

【0030】入力手段133はテンキー、キーボード等の文字入力装置を使用する。入力手段133からは特徴収集モードおよび音声認識モードの指示、収集した特徴に付加して記憶する識別情報、たとえば、識別番号、説明文等、および音声認識に使用する特徴をケプストラム平均格納手段166から取り出すときに与える識別情報を入力する。

【0031】本実施の形態では、複数の話者のケプストラム平均や、異なるマイクロホンのケプストラム平均をケプストラム平均格納手段166に登録しておき、音声認識時に歪み除去のために使用するケプストラム平均を選択することに特徴がある。

【0032】音声ケプストラム平均計算手段141はデジタル音声信号SS中の音声区間(図2参照)についてのケプストラム平均を次式により計算する。

【0033】

$$【数1】VC^{new}(i) = \alpha \cdot VC^{old}(i) + (1 - \alpha) \cdot C_t(i)$$

ここで $C_t(i)$ はフレーム番号tにおけるi次のケプ

ストラムである。VC^{new} は今回計算すべきケブストラム平均、VC^{old} は前回計算したケブストラム平均である。 α は固定の係数である。

【0034】なお、VC^{old} の初期値として、音声認識時には、ケブストラム平均格納手段166に予め登録され、指定されたケブストラム平均が使用される。特定話者の特徴およびある環境下での音響の特徴の最初の収集時には指定されたケブストラム平均が存在しないため初期値として数値0が使用される。計算結果を音声ケブストラム平均VCと表す。

【0035】非音声ケブストラム平均計算手段142も上記数1式を使用してケブストラム平均を計算する。初期値の取り方も上述と同様である。計算結果を非音声ケブストラム平均UCと表す。

【0036】音声ケブストラム減算手段151は入力ケブストラムCから音声ケブストラム平均VCを減算する。これによりマイクロホンの周波数特性の相違、話者の声道、声帯の違いによる音声特性の相違や話者とマイクロホンの位置関係の相違に起因するスペクトル領域の乗法性歪みを一括して補正する。

【0037】非音声ケブストラム減算手段152は入力ケブストラムCから非音声ケブストラム平均UCを減算する。これによりマイクロホンなどの音声収録の信号処理系の相違により生じるスペクトル領域の乗法性歪みを補正する。音声区間の歪み補正後のケブストラム平均および非音声区間の歪み補正後のケブストラム平均が照合手段188に送られる。なお、音声ケブストラム平均計算手段141と非音声ケブストラム平均計算手段142の切り換えおよび音声ケブストラム減算手段151と非音声ケブストラム減算手段152の切り換えは判定手段122の判定計結果（フラグVAD）に応じて行われる。

【0038】読み出し手段171は音声認識時、入力手段から指示されたケブストラム平均をケブストラム平均格納手段166から読み出し、音声ケブストラム平均計算手段141または非音声ケブストラム平均計算手段142に初期値として供給する。また、読み出し手段171は音声ケブストラム平均計算手段141、非音声ケブストラム平均計算手段142がそれぞれケブストラム平均を計算しているときには、前の時点のケブストラム平均VC^{old} を読み出して対応のケブストラム平均計算手段141、142に供給する。

【0039】書き込み手段172は音声、音響特徴を収集時には音声ケブストラム平均計算手段141および非音声ケブストラム平均計算手段142の計算結果をケブストラム平均格納手段166に保存目的で格納する。書き込み手段172は音声認識時には音声ケブストラム平均計算手段141および非音声ケブストラム平均計算手段142の計算結果を次の計算のためにケブストラム平均格納手段166に一時格納する。

【0040】ケブストラム平均格納手段166は登録用のケブストラム平均の保存領域とケブストラム平均の計算用の一時格納領域を有する。登録用のケブストラム平均記憶領域にはケブストラム平均と識別情報が互いに関連付けられて記憶され、入力手段133からの指示で読み出し手段により全ての識別情報を読み出して不図示の表示手段に表示することが可能である。これによりユーザは、ケブストラム平均格納手段に登録されているケブストラム平均の種類内容を知ることができる。ケブストラム平均の記憶領域は音声用と非音声用の2種類が用意されている。

【0041】照合手段188は従来と同様、動的計画法や隠れマルコフモデル法により補正後のケブストラム平均NCを使用して音声認識を行う。

【0042】このようなシステム構成の動作説明を行う。

【0043】(A) 音声・音響環境特徴の収集
ユーザが自分の声の特徴を登録する場合に、入力装置133から音声収集に関する識別情報を入力手段133から入力する。次に、マイクロホンから音声を入力する。マイクロホンからのアナログ音声信号Sはデジタル音声信号SSに変換される。音響分析手段111によりケブストラム分析がフレーム単位で行われる。判定手段122ではデジタル音声信号SSが音声か非音声かをフレーム単位で判定する。音声の判定が得られた場合、音響分析手段111の接続先の処理系が音声ケブストラム平均計算手段141に設定される。

【0044】これにより、音声ケブストラム平均計算手段141は、ケブストラム平均の初期値として数値0を設定した後、音響分析手段111から出力された入力ケブストラム平均と初期値との平均を数1式により計算し、その計算結果を書き込み手段172に引き渡す。書き込み手段172は上記計算結果を次のケブストラム平均の計算のためにケブストラム平均格納手段166の中の一時記憶領域に記憶する。

【0045】音声が入力されている間、上述のケブストラム平均の更新計算が繰り返し行われる。非音声の判定が得られると、その時点で書き込み手段172は音声ケブストラム平均計算手段141の計算結果をケブストラム平均格納手段166の保存領域に新規登録する。このとき、音声入力に先立って、入力手段133から入力された識別情報をも上記保存領域に登録する。

【0046】以上が話者の音声特徴、話者とマイクロホンの位置関係、マイクロホンの特性の3つの総合的なスペクトル特性（乗法性歪み）を収集する例であるが、マイクロホンなどの信号処理系のスペクトル特性（乗法性歪み）については、音声を入力しない状態で、周囲の音響のみをマイクロホンにより入力する。判定手段122の判定結果が非音声となるので、非音声ケブストラム平均計算手段142が選択され、上述と同様にして、非音

声ケブストラム平均の計算結果をケブストラム平均格納手段166に格納することができる。なお、計算の終了の指示方法としては、ユーザが音声を入力して、判定手段122の判定が非音声から音声に切り変わった時点を終了時点としてもよいし、入力手段133から終了の指示を行ってもよい。非音声のケブストラム平均と共に識別情報がケブストラム平均格納手段166の保存領域に格納されることは言うまでもない。

【0047】(B) 音声認識

音声認識に先立って、ユーザはケブストラム平均格納手段166の保存領域に保存されているケブストラム平均を指定する。話者の異なり、話者とマイクロホンの位置関係の異なりおよびマイクロホンなどの収録環境の異なりに応じた音声のケブストラム平均とマイクロホンなどの収録環境の異なりに応じた非音声のケブストラム平均を指定する。

【0048】これにより、指定されたケブストラム平均がケブストラム平均格納手段166から読み出し手段171により読み出され、識別情報の示す音声/非音声に応じて対応のケブストラム平均計算手段141、142に初期値として供給される。ケブストラム平均が指定されなかった場合、または指定しても対応するケブストラム平均が格納手段166に存在しない場合には初期値0が与えられる。

【0049】マイクロホンから入力されたアナログ音声信号Sは変換手段100によりデジタル音声信号SSに変換され、音響分析手段111によりケブストラム分析される。デジタル音声信号SSが音声区間の間は音声ケブストラム平均計算手段141が音声ケブストラム平均計算手段141、音声ケブストラム減算手段151の信号処理系が選択されて、照合手段188に歪み補正したケブストラムNCが供給される。

【0050】デジタル音声信号SSが非音声区間の間は非音声ケブストラム平均計算手段142、非音声ケブストラム減算手段152の信号処理系が選択されて、歪み補正したケブストラムNCが供給される。

【0051】(第2の実施形態)第1の実施形態は音声、音響の特徴収集機能と、歪み補正機能を有する音声認識装置の例であったが、双方の機能を別体の装置で構成する第2の実施形態を説明する。図3は音声、音響特徴収集用装置のシステム構成を示し、図4は歪み補正機能を有する音声認識装置のシステム構成を示す。

【0052】図3、図4において、図1と同様の個所については同様の個所には同一の符号を付しており、第1の実施形態との相違点を説明する。図3に示す装置では、読み出し手段171、音声ケブストラム減算手段151、非音声ケブストラム減算手段152、照合手段188を有しておらず、ケブストラム平均格納手段166に収集した音声、音響の特徴を識別情報を付して登録することのみが可能である。

【0053】ケブストラム平均格納手段166には、ICカード、メモ리카ード、フロッピーディスクのような携帯用記憶媒体を使用する。ユーザはこの装置を携帯し、任意の場所で音響や音声の特徴を収集できる。

【0054】このようにして、収集した音声、音響の特徴を記憶したケブストラム平均格納手段166を図4の装置に装着して、入力音声の歪み補正を行う。

【0055】図4の装置では、ケブストラム平均格納手段166から読み出されたケブストラム平均が音声ケブストラム減算手段151および非音声ケブストラム減算手段152に直接送られ、入力ケブストラムCの減算に使用される。

【0056】(C) 実施形態3

第1の実施形態の具体的なシステム構成を図5に示す。図5において、音声認識装置は、CPU200、ROM210、RAM220、A/D変換器230、メモ리카ード用アダプタ240、キーボード250およびLCD(液晶表示器)260を有する。CPU200はROM210に格納された図6の処理プログラムを実行して、音声、音響の特徴を収集すると共に、入力音声の歪み補正、音声認識を行う。ROM210は図6の処理プログラムを記憶するとともに、周知のシステムプログラム、およびシステムプログラムで使用するデータ、たとえば、フォントデータ等を記憶する。RAM220はCPU200に入出力するデータを一時記憶する。RAM220が図1のケブストラム平均格納手段166のケブストラム平均の一時記憶領域として機能する。

【0057】A/D変換器230はマイクロホンと接続し、変換手段100としてマイクロホンから入力されたアナログ音声信号Sをデジタル音声信号SSに変換してCPU200に引き渡す。アダプタ240はメモ리카ード245と脱着可能に接続する。メモ리카ード245は、ケブストラム平均格納手段166の保存領域として機能し、収集したケブストラム平均を登録するとともに、音声認識のための標準パターンを記憶している。メモ리카ード245に対する読み書きはCPU200の制御で行われる。キーボード250は入力手段133として機能する。LCD260は音声認識結果を表示すると共に、メモ리카ード245に格納されているケブストラム平均の識別情報(識別番号、説明)を表示する。

【0058】このようなシステム構成の動作を図6のフローチャートを参照して説明する。なお、図6に示す処理手順は、CPU200により読み取り実行可能なプログラム言語で記載されているが、説明の便宜上、機能表現している。

【0059】図6において、ユーザはキーボード250から処理目的に合致したモードを指示する。この形態では、音声認識モード、音声、音響特徴収集モード、再生モードが用意されている。音声認識モードおよび使用するメモ리카ード245内のケブストラム平均が指示され

た場合には、手順はステップS10→S20→S200へと進み、指定されたケプストラム平均を読み出した後、音声の入力、歪み補正のための信号処理（図1の音響分析手段111、判定手段122、音声ケプストラム平均計算手段141、音声ケプストラム減算手段151およびまたは非音声音声ケプストラム平均計算手段142手段ならびに非音声ケプストラム減算手段152、読み出し手段171により実行される処理）、歪み補正後のケプストラム平均を使用した音声認識処理（図1の照合手段188の処理）がCPU200において音声入力の間繰り返して行われる（ステップS200～S230のループ処理）。

【0060】音声認識の結果得られる文字列は、LCD260に表示される。

【0061】一方、音声、音響特徴収集モードが指示されると、手順はステップS10→S20→S300へと進み、キーボード250からのスタートの指示を受けて、入力音声／非音声についてのケプストラム平均を取得する（ステップS310→S330のループ処理）。この処理を実行する間、CPU200が音響分析手段111、音声ケプストラム平均計算手段141および／または非音声ケプストラム平均計算手段142、判定手段122、書き込み手段172として動作する。キーボード250からの終了の指示に応じて、CPU200はキーボード250から識別番号、音声／非音声の種類情報、データ収集環境の説明を入力する。最後にCPU200はRAM220上に一時記憶したケプストラム平均とキーボード250から入力された上記情報を一対として、メモカード245に書き込む（ステップS340→S350）。

【0062】

【実施例】第1の実施形態を適用して隠れマルコフモデル法により不特定話者の音声認識を行った結果、従来のケプストラム平均正規化方法に比べて2%の性能向上が見られた。この音声認識実験では、マイクロホンを運転席サンバイザに設置し、運転席に座った男女2名ずつ計4名の各話者がそれぞれ発声した520単語の音声データ計2080データ（データ1）および助手席に座った男女2名計4名の話者がそれぞれ発声した520単語の音声データ計2080データ（データ2）を収録し、評価データとして使用した。このとき、自動車のエンジンは停止している。

【0063】従来のケプストラム正規化法では、データ1、データ2の認識率は90.2%、87.7%であった。これに対し、第1の実施形態を適用した場合には、データ1、データ2の認識率は92.7%、89.5%であり、ともに2%の性能向上が認められた。

【0064】上述した実施形態の他に次の形態を実施できる。

【0065】1）上述の実施形態では、予め複数の話者

の音声や音響環境の音響の歪みの特徴を登録し、登録の特徴に基づき補正を行って、音声認識を行っている。

【0066】たとえば、複数の人間に携帯用記憶媒体をそれぞれ配布し、各人間が自分の音声特徴と、音声認識装置を使用する際の音響環境特徴を登録しておく。このようにして、1台の音声認識装置を複数の人が共有することができる。この場合には、携帯用記憶媒体を音声認識装置に装着し、携帯用記憶媒体から歪み補正に使用する特徴データを読み出し、音声認識時に使用する。

【0067】2）音声認識装置の記憶手段に記憶した特徴データの取り出しには以下のような方法がある。第1には、マイクロホンの距離が遠い、近いの2段階に分け、予め特定環境下で特徴データを収録する。音声認識時は上記遠い、近いの段階を指示し、対応する特徴データを取得する。人間の音声特徴（歪み）を記憶する場合には氏名などを付して記憶し、氏名により対応の音声特徴を取り出すとよい。同一の人間でも、早口、通常の2段階に分けて音声特徴を記憶し、段階の指示により対応の音声特徴を取り出すとよい。

【0068】3）ケプストラム平均は数1式に限らず、相加平均等の平均式を使用することができるが、たとえば、音声の場合、音声区間が終了しないと平均計算ができないので、処理時間を短縮したい場合には数1式のほうが好ましい。

【0069】

【発明の効果】請求項1、10の発明では、特定話者の音声特徴や音響等の複数種の音の特徴をユーザ自身が登録でき、かつ、選択使用して音声認識時の音声の歪み補正を行うので、たとえば、マイクロホンと話者の口との間の距離の相違、話者の相違、特定話者の話す速度の相違等に起因する誤認識を解消できる。

【0070】請求項2、11の発明では、音の特徴をケプストラム分析により取得する。ケプストラム分析では、入力信号のフーリエ変換の対数を取得し、その逆フーリエ変換を行うので、たとえば、マイクロホンの歪み特性、音声の歪み特性、壁などにより反射されてマイクロホンに入力される音響の歪みの積はケプストラム領域において各歪み特性に対するケプストラムの和として得られるという利点がある。この利点を使用することにより予め収集した特定の種類の音の（話者の相違、マイクロホンの相違等）の特徴を使用することで、音声認識時に入力信号に含まれる複数種類の音の中の上記特定の種類の音に対してその歪み補正を簡単に行うことが可能となる。

【0071】請求項3、4、12、13の発明では、収集した音の特徴を、付加された情報に基づきユーザが取り出すことができ、また、情報を表示することにより、音の特徴の内容、たとえば、話者の名前、音響の収録環境等を知ることができる。さらに記憶してある特徴の中からユーザは好適な音の特徴を選択することができる。

【0072】請求項5、6、14、15の発明では、音声と、音響の種類判別ができ、特定話者や音響環境等の多種の信号歪みに対処することができる。

【0073】請求項7、16の発明では、音の特徴を脱着可能な記録媒体に記憶し、各ユーザが記録媒体を保有することで、音声認識装置を共有使用することができる。

【0074】請求項8、17の発明では、音声／非音声を自動識別することにより入力音声の中から音声／非音声を抽出することができる。

【0075】請求項9、18の発明では、音の特徴の収集機能はなく、予め製造者側等で用意した音の特徴をユーザが選択して使用する。これにより、たとえば、マイクロホンの設置位置と話者の口との距離の遠近等をユーザが指定するだけで、好適な音声認識結果が得られ、音の特徴データの収集操作からユーザは開放される。また、製造者側では、マイクロホンの種類毎に音のデータを記憶しておけばよいので、たとえば、会議場で話者から離れた位置にマイクロホンが設置された場合での音声認識装置、話者の傍にマイクロホンが設置された場合での音声認識装置というように多種の音声認識装置を提供できる。また、装置のハード自体の仕様変更がないので、製造コストに影響を与えることはない。

【図面の簡単な説明】

【図1】本発明第1の実施形態のシステム構成を示すブ

ロック図である。

【図2】音声／非音声の判定処理を説明するための説明図である。

【図3】第2の実施形態の補正用データ収集装置のシステム構成を示すブロック図である。

【図4】第2の実施形態の音声認識装置のシステム構成を示すブロック図である。

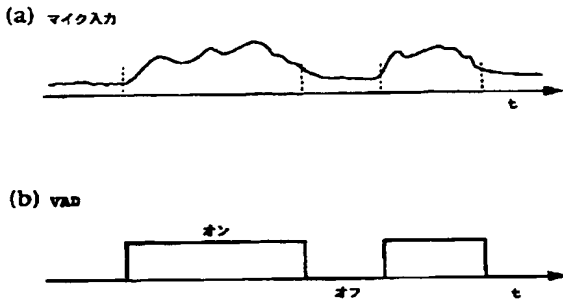
【図5】実施形態3のシステム構成を示すブロック図である。

【図6】実施形態3の処理手順を示すフローチャートである。

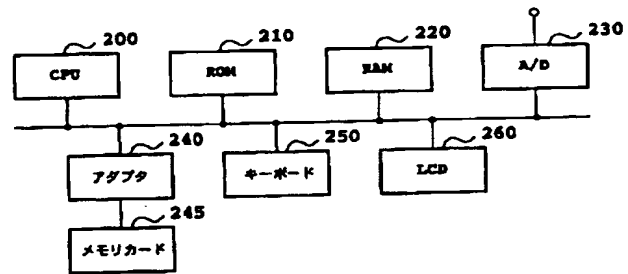
【符号の説明】

- 100 変換手段
- 111 音響分析手段
- 122 判定手段
- 133 入力手段
- 141 音声ケプストラム平均計算手段
- 142 非音声ケプストラム平均計算手段
- 151 音声ケプストラム減算手段
- 152 非音声ケプストラム減算手段
- 166 ケプストラム平均格納手段
- 171 読み出し手段
- 172 書き込み手段
- 188 照合手段

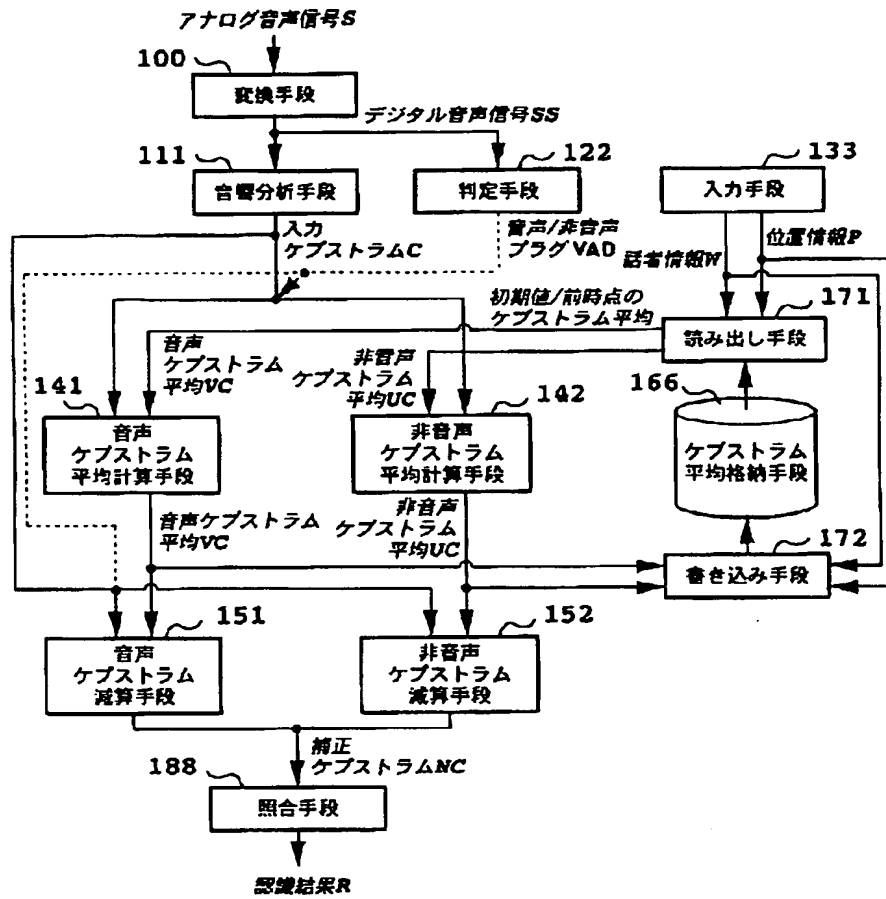
【図2】



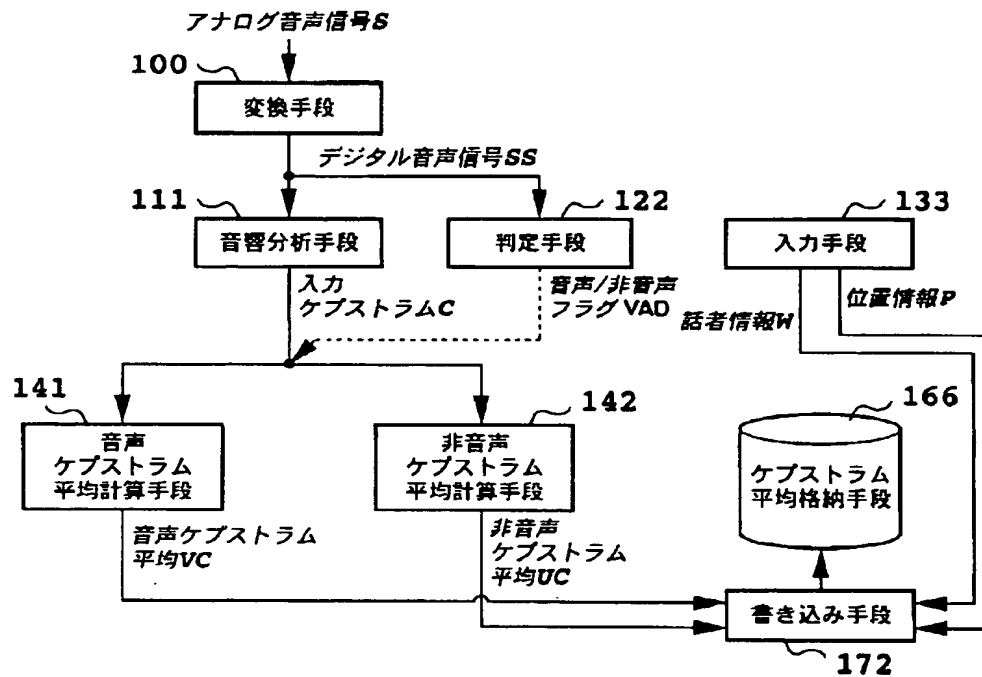
【図5】



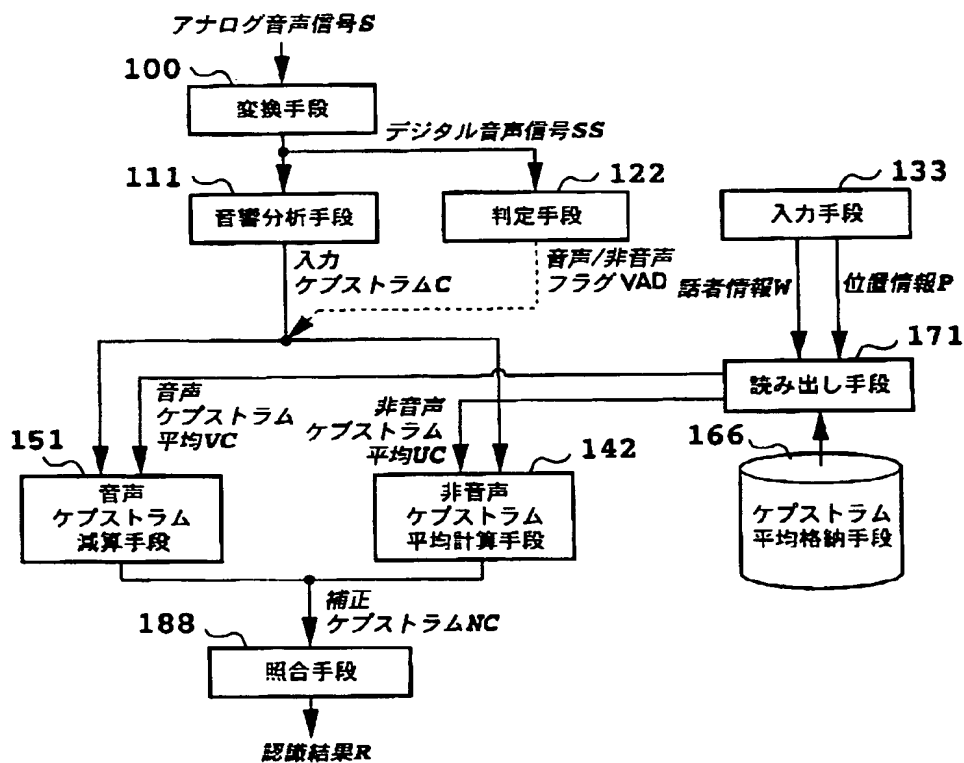
【図1】



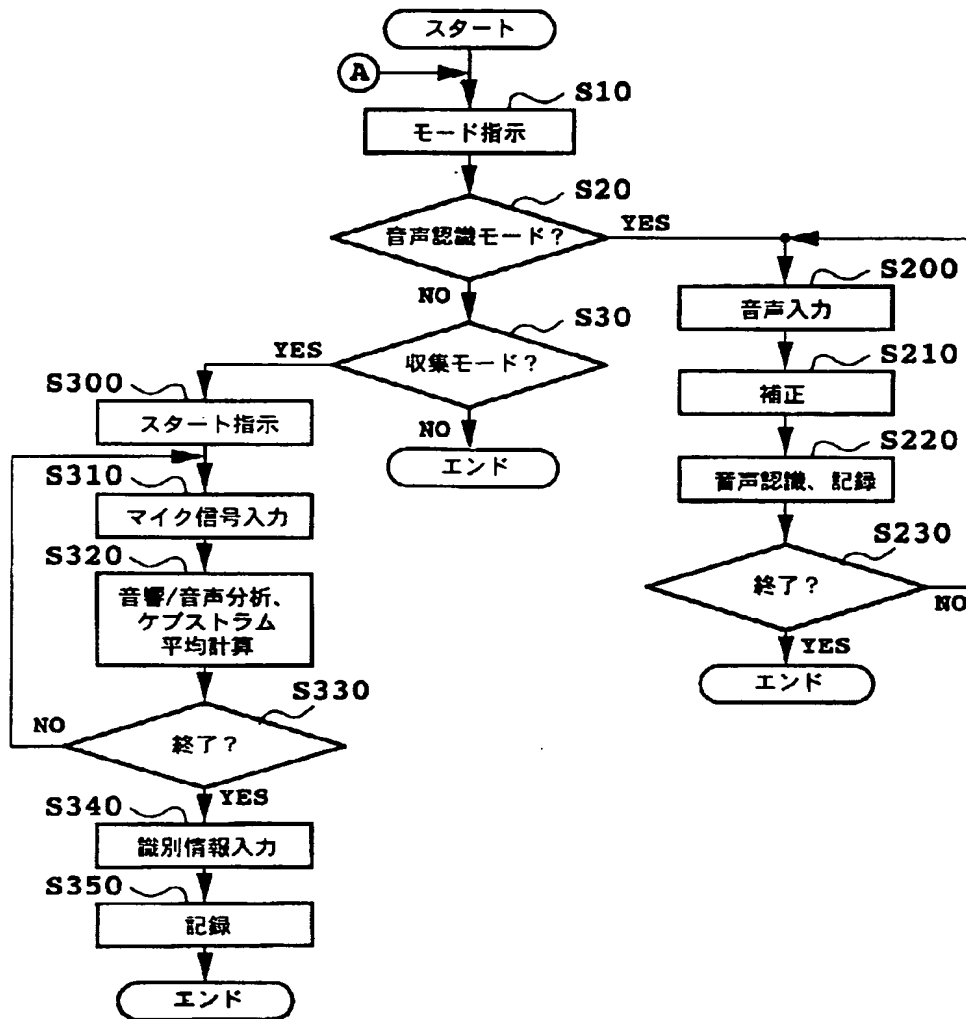
【図3】



【図4】



【図6】



**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☒ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.